

文章编号:1671-6833(2025)06-0049-09

基于 CSI 主成分分割的人体动作识别方法

饶 壮¹, 丁大钊², 王依菁²

(1. 郑州大学 网络空间安全学院, 河南 郑州 450002; 2. 嵩山实验室, 河南 郑州 450000)

摘要:传统的基于信道状态信息(CSI)进行人体动作识别的方法具有输入数据冗余、提取特征单一等问题。基于此,提出了一种基于CSI主成分的双层滑动窗口机制分割的人体动作识别方法。首先,对振幅进行去异常值、降噪,对相位进行线性校准、降噪;其次,利用基于主成分分析的双层滑动窗口机制对预处理后的CSI数据进行活动分割,去除与运动无关的信息,提升模型的训练效率;再次,通过卷积神经网络与双向循环门控单元对CSI数据的空间和时间两个维度进行分析,并融合多头注意力机制聚焦关键信息,实现对人体动作的高精度识别;最后,在WiAR和BAHAR两个公开数据集上进行实验验证,结果表明:所提方法可以有效地对多种环境下的多种人体活动进行识别,并减少50%的数据量,WiAR数据集准确率达96.53%,极大改善了输入数据冗余和提取特征单一问题,所提方法优于其他现有的识别方法。

关键词:信道状态信息;活动分割;卷积神经网络;双向门控循环单元;多头注意力机制

中图分类号: TN92; TP183; TP391.4 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2025.03.021

随着互联网和Wi-Fi设备的大规模普及,复用这些无处不在的Wi-Fi信号,在自然状态下以无接触方式对人体信息进行感知,已成为一个新兴且极具潜力的研究方向。研究者在小尺度的呼吸监测^[1]、手势识别^[2]、唇语识别^[3],到大尺度的动作识别^[4]、入侵检测^[5]、人数统计^[6]等方面都进行了大量尝试。尤其是动作识别,在健康监测^[7]、虚拟现实^[8]、智能家居^[9]等实际应用中发挥重要作用。因此,提高利用Wi-Fi信号对人体动作进行识别的准确率具有重要意义。

目前,对人体动作进行识别的技术主要有3种:第1种是基于计算机视觉的动作识别技术,该技术通过相机采集图像或视频信息,对采集环境的光线和相机的像素、拍摄范围有着较高的要求,且有着泄露隐私的风险;第2种是基于可穿戴传感器的动作识别技术,由于其需要随时佩戴,对于一些如老人、小孩等特殊人群部署有较大的困难,且长时间佩戴易造成不便;第3种是基于射频信号的动作识别技术,包括Wi-Fi信号和其他特制射频信号,由于特制射频信号用于特定应用场景或具有特定功能,所以

存在成本高昂、调制和维护困难等缺陷,不利于大规模应用。利用Wi-Fi信号进行人体动作识别凭借其成本低、隐私安全性高和易部署等特点,逐渐受到研究者的广泛关注。

基于Wi-Fi信号的无线感知主要分为两类:基于接收信号强度(received signal strength indication, RSSI)^[10]和基于信道状态信息(channel state information, CSI)。由于RSSI是多径信号叠加的结果,因此只能用于实现一些粗粒度的感知任务。随着Halperin团队在2011年通过修改Intel 5300 NIC驱动开发出了首款对商用Wi-Fi网卡采集CSI的专业工具Linux 802.11n CSI Tool^[11],研究者们才逐渐选择以CSI作为分析的对象。与基于RSSI仅考虑叠加功率强度的感知对象不同,基于CSI的感知可以提供更多细粒度的信号变化特征,如组成完整信号的每个子载波的振幅与相位信息,从而具有更加精细识别人体动作的能力。

使用基于特征的机器学习算法可以有效地处理人体动作的识别问题, Li等^[12]对CSI数据手动提取特征,并使用随机森林算法进行动作分类; Guo

收稿日期:2025-01-30; 修订日期:2025-03-21

基金项目:河南省科技攻关计划项目(232102210045);嵩山实验室重大科研项目(ZZK202403002)

通信作者:丁大钊(1979—),男,河南许昌人,嵩山实验室副研究员,主要从事无线网络安全及无线通信感知一体化研究, E-mail:1007786094@qq.com。

引用本文:饶壮,丁大钊,王依菁. 基于CSI主成分分割的人体动作识别方法[J]. 郑州大学学报(工学版), 2025, 46(6): 49-57. (RAO Z, DING D Z, WANG Y J. Human activity recognition method based on CSI principal component segmentation[J]. Journal of Zhengzhou University (Engineering Science), 2025, 46(6): 49-57.)

等^[13]对比了多种机器学习算法在同一人体动作数据集的分类识别情况,发现不同的算法擅长识别不同的动作类别,且深度学习算法平均准确率高于机器学习算法。

深度学习一些常见模型包括卷积神经网络(CNN)、循环神经网络(RNN)等,其中CNN在空间频域分析^[14]、RNN在时间序列分析^[15]中具有良好效果。Moshiri等^[16]将CSI的振幅在 $[0, 255]$ 上进行归一化将其转换为伪彩色图,并将图像输入到二维卷积神经网络,准确率达到95%以上;Shi等^[17]过滤掉了与活动无关的CSI信息,通过长短期记忆循环神经网络提取时序特征进行人体动作识别;Shang等^[18]将CSI的振幅去噪平滑处理后输入到CNN和LSTM的组合模型中,取得了94.14%的准确率。这类方法要么单独使用CNN或RNN及其变体,要么将CNN与其他模型相结合^[16-18]。尽管RNN等模型可以在时间序列分析上取得好的效果,但此类分析仅面向过去时间,忽视了对未来时间的考量。在上述已有方案中,研究者们或将采集到的所有CSI信息输入到神经网络中,使得特征选择和特征工程变得更加困难,导致模型的性能下降,进而影响模型的预测准确性;或利用简单的滑动窗口机制去除掉与运动无关的CSI信息,由于受多径效应影响,信号会产生不规则波动,简单的滑动窗口机制无法有效筛选掉无用的信息,甚至存在误将有用信息去除的可能,使得模型更难学习有效的特征。不仅如此,Yamak等^[19]发现循环门控单元(gated recurrent unit, GRU)在时间序列预测中比LSTM模型表现更出色。此外,上述大多方案仅将CSI的振幅信息作为分析的起点,而忽略了相位信息的重要性,Ming等^[20]发现综合考虑振幅与相位的情况的精度最高。

针对以上问题,本文提出了一种基于CSI主成分的双层滑动窗口机制的人体动作识别方法,并采用公开数据集进行测试。本文的主要贡献为①基于主成分分析,利用双层滑动窗口机制提出了一种能够提取CSI有效运动信息的方法,去除掉与运动无关的信息,以此提升模型的训练效率;②提出了一种基于CNN-BiGRU的架构,对CSI数据的空间和时间两个维度进行分析,并融合多头注意力(multi-head attention, MHA)机制,聚焦关键信息部分,从而提高识别精度;③综合考量振幅和相位对识别精度的影响,将相位作为输入模型的数据之一,通过实验验证相位对提升动作识别精度的重要性。

1 系统模型与识别流程

所提系统流程如图1所示,整个系统由数据预处理、活动分割和动作分类三部分构成。首先,对原始CSI进行预处理,这一过程包括对振幅进行去异常值、降噪处理,以及将相位线性化并降噪;其次,依次对3个收发天线对的振幅开展主成分分析,在此基础上运用双层滑动窗口机制,从而得到包含有效运动信息的窗口索引;最后,依据该索引对预处理后的振幅与相位筛选无关信息,并通过神经网络实现对人体动作的识别。

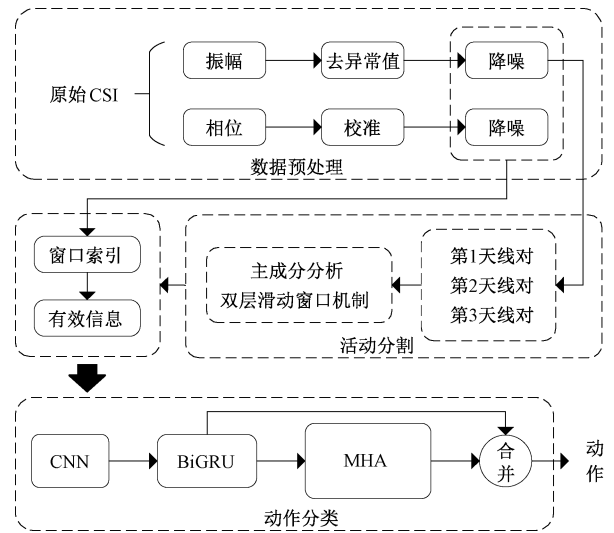


图1 系统流程图

Figure 1 System flow chart

1.1 信道状态信息

在频域中,无线信道可以被描述为

$$\mathbf{R} = \mathbf{H} \times \mathbf{T} + \mathbf{N}. \quad (1)$$

式中: \mathbf{H} 表示CSI的信道矩阵; \mathbf{R} 和 \mathbf{T} 分别为接收方和发射方的信号向量; \mathbf{N} 为信号传播过程中产生的噪声向量。假设系统的接收方具有 p 根天线,发送方具有 q 根天线, \mathbf{H} 可以表示为

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,r} & \cdots & \mathbf{H}_{1,p} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,r} & \cdots & \mathbf{H}_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{H}_{t,1} & \mathbf{H}_{t,2} & \cdots & \mathbf{H}_{t,r} & \cdots & \mathbf{H}_{t,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{H}_{q,1} & \mathbf{H}_{q,2} & \cdots & \mathbf{H}_{q,r} & \cdots & \mathbf{H}_{q,p} \end{bmatrix}. \quad (2)$$

式中: $t \in \{1, 2, \dots, q\}$ 且 $r \in \{1, 2, \dots, p\}$ 。每个接收到的CSI数据包都能解析出 $p \cdot q \cdot n$ 个子载波,其中 n 表示每个无线信道中子载波的数量。本文采用的公开数据集均为利用Linux 802.11n CSI Tool从Intel 5300 NIC中采集的数据, $p = 1, q = 3, n = 30$,则每个收发天线对有

$$\mathbf{H}_{i,r} = [H_{i,r}^1, H_{i,r}^2, \dots, H_{i,r}^{30}]. \quad (3)$$

且每个子载波 $H_{i,r}^i$ 有

$$H_{i,r}^i = H_{i,r}^i e^{j \sin \angle H_{i,r}^i}, i \in [1, 2, \dots, 30]. \quad (4)$$

式中: $H_{i,r}^i$ 和 $\angle H_{i,r}^i$ 分别为第 i 个子载波的振幅和相位。由此可见,对信道矩阵 \mathbf{H} 进行分析,就能进一步提取出子载波的振幅和相位,进而实现对人体动作的识别。

1.2 数据预处理

1.2.1 异常值去除

由于 CSI 具有较高的敏感性,受到硬件或环境等各类不确定因素影响时会产生异常值,考虑到这些异常值会对后续分析造成干扰,因此有必要对其进行处理。Hampel 滤波是一种有效的异常值去除方法,其基本原理是利用移动平均窗口查找窗口内超出 $[\mu - \eta\sigma, \mu + \eta\sigma]$ 的异常值并将其替换为 μ , 其中 μ, σ 分别为窗口中数据的中值和平均绝对偏差, η 为调整参数,本文选取默认值 3。对 CSI 的振幅进行了异常值去除处理,效果如图 2 所示。

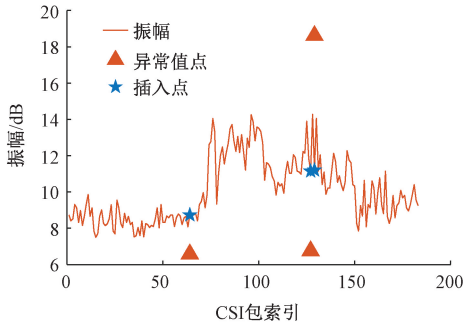


图 2 Hampel 去除异常值

Figure 2 Hampel removes outliers

1.2.2 相位校准

为有效消除采集到的原始 CSI 相位信息中存在的频率偏移与时钟偏移,本文采用解卷绕和线性变换的方法对相位进行处理。

由于相位信号是周期性的,会出现相位跳跃现象,进而导致不连续性。解卷绕的目的是消除这些跳跃,确保相位变化平滑且连续。解卷绕通过分析相位值的增减变化来确定信号的真实相位角度,在对原始相位进行解卷绕后,第 i 个子载波的测量相位 $\hat{\phi}_i$ 可以表示为

$$\hat{\phi}_i = \phi_i - 2\pi \frac{k_i}{B} \Delta t + \beta + Z. \quad (5)$$

式中: ϕ_i 为真实相位; k_i 为第 i 个子载波索引; B 为快速傅里叶变换的长度; Δt 为接收方时钟误差; β 为未知相位偏移; Z 为测量过程中的噪声,可忽略不计。为了消除 Δt 和 β 的影响,引入 a, b 这两个线

性相关量以进行线性变换:

$$a = \frac{\hat{\phi}_n - \hat{\phi}_1}{k_n - k_1} = \frac{\phi_n - \phi_1}{k_n - k_1} - \frac{2\pi}{B} \Delta t; \quad (6)$$

$$b = \frac{1}{2}(\hat{\phi}_1 + \hat{\phi}_n) = \frac{1}{2}(\phi_1 + \phi_n) - \frac{\pi \Delta t}{B}(k_1 + k_n) + \beta. \quad (7)$$

根据 IEEE 802.11 n 协议,子载波索引 k_1 与 k_n 的取值是完全正负对称的,则 $\frac{\pi \Delta t}{B}(k_1 + k_n) = 0$, 因此 $b = \frac{1}{2}(\phi_1 + \phi_n) + \beta$ 。通过 $\hat{\phi}_i$ 减去 $ak_i + b$, 线性

变换后的校准相位 $\tilde{\phi}_i$ 可表示为

$$\tilde{\phi}_i = \hat{\phi}_i - ak_i - b = \phi_i - \frac{\phi_n - \phi_1}{k_n - k_1} k_i - \frac{1}{2}(\phi_1 + \phi_n). \quad (8)$$

其中, $\frac{\phi_n - \phi_1}{k_n - k_1}$ 与 $\frac{1}{2}(\phi_1 + \phi_n)$ 是固定值, k_i 已知。

通过上述处理,可得到与真实相位线性相关的校准相位,以此来替代真实相位进行分析,效果如图 3 所示。

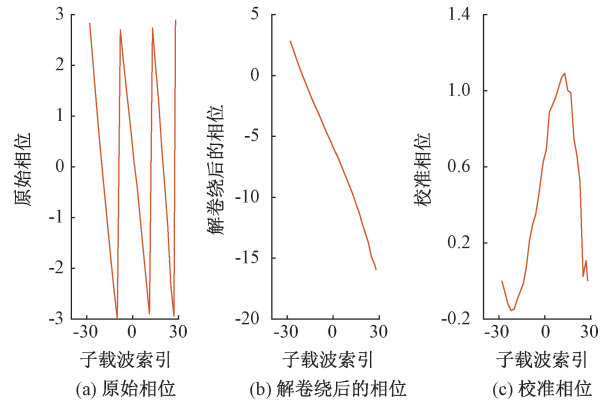


图 3 相位线性变换

Figure 3 Phase linearization

1.2.3 降噪

对振幅去除异常值和相位进行线性变换后,还存在很多噪声,本文使用 Butterworth 低通滤波器分别对其高频噪声进行降噪,如图 4 所示。具体来说,低通滤波器可以让频率低于设定阈值的信号顺利通过,但是会阻挡和减弱高于这个阈值的信号,由于人体动作多保留在 CSI 的低频部分,所以低通滤波器能够有效滤除与人体动作无关的噪声。

1.3 活动分割

本文先对振幅进行降维,得到第一主成分,由于非活动部分的振幅变化程度低于活动部分,在第一

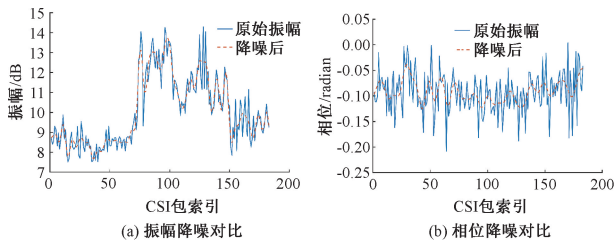


图4 振幅与相位降噪对比

Figure 4 Contrast of amplitude and phase noise reduction

主成分上应用双层滑动窗口机制,筛除非运动信息,流程如图5所示。

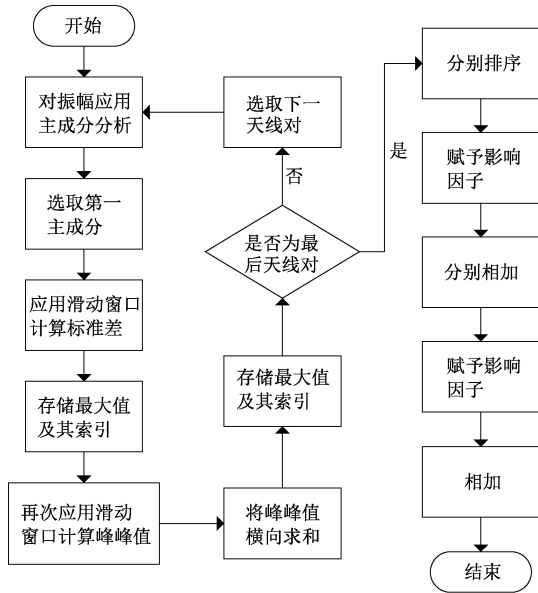


图5 活动分割算法流程图

Figure 5 Activity segmentation algorithm flow chart

算法1 基于主成分分析的双层滑动窗口机制的活动分割算法。

输入:CSI的3个收发天线对的振幅。

输出:运动信息的窗口索引。

- ① 循环,依次选取3个天线对
- ② for $i = 1 : 3$
- ③ 对第 i 个天线对的振幅进行主成分分析;
- ④ 提取该天线对振幅主成分分析后的第一主成分;
- ⑤ 对第一主成分应用窗口大小为 m 的滑动窗口,计算每个窗口的标准差,得到数组 $std(w,1)$, w 为窗口个数;
- ⑥ 将最大的标准差及其索引存储到数组 $L(i,1)$ 和 $L(i,2)$ 中;
- ⑦ 对⑤中的每个窗口再次应用窗口大小为 $n(m > n)$ 的滑动窗口,计算每个窗口的峰峰值,得到数组 $ptp(w,s)$, s 为二次窗口个数;
- ⑧ 对 ptp 进行横向求和;
- ⑨ 将最大的峰峰值横向和及其索引存储到数组

$L(i,3)$ 和 $L(i,4)$ 中;

⑩ end

⑪ 对 L 存储的标准差和峰峰值排序,分别对其最大值、中值、最小值对应的索引赋予影响因子,本文选取 0.5, 0.3 和 0.2, 相加得到综合 3 个天线对振幅变化程度最强烈的标准差和峰峰值对应的索引;

⑫ 对标准差和峰峰值对应的索引再次赋予影响因子,本文选取 0.3 与 0.7, 最终相加计算出运动信息的窗口索引。

算法考虑了 CSI 的 3 个收发天线对由于位置和对信号感知能力不同的特性,对每个天线对设计了不同的影响因子。标准差作为衡量振幅各个数据点与振幅平均值之间的差异程度的指标,可捕捉人体动作整体的活动时长信息;峰峰值作为衡量振幅范围的指标,当窗口大小设置恰当时,一定程度上能够捕捉人体动作开始和结束的时刻信息,对活动分割具有重要意义。活动分割算法效果如图6所示。

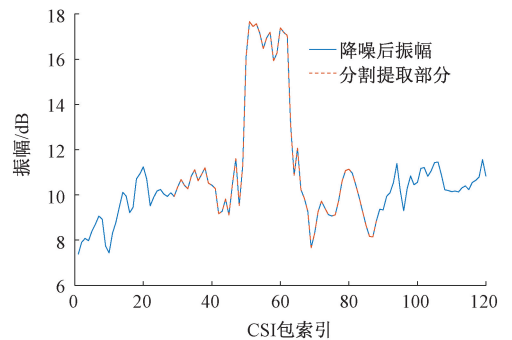


图6 活动分割算法效果图

Figure 6 Activity segmentation algorithm effect

1.4 CNN-BiGRU 模型

为全面挖掘数据潜在特征间的关系,对 CSI 数据的空间和时间两个维度进行综合剖析,并兼顾过去时间与未来时间对人体动作识别的影响,本文采用 CNN-BiGRU 的神经网络模型架构。其中,CNN 部分借助卷积层和池化层提取局部特征,并通过激活函数增强非线性能力;BiGRU 部分运用双向结构和门控机制捕捉数据的前后依赖关系,以处理时序特征。此外,为使模型聚焦于关键信息,融入多头注意力机制,进而提升模型性能。

模型的输入是一个单通道的二维矩阵,行向量由 3 个收发天线对的振幅和相位拼接而成,根据式(3)和式(4),行向量可以表示为

$$\vec{H} = [\mathbf{H}_{1,1}, \mathbf{H}_{1,2}, \mathbf{H}_{1,3}, \angle \mathbf{H}_{1,1}, \angle \mathbf{H}_{1,2}, \angle \mathbf{H}_{1,3}]。 \quad (9)$$

即行向量 \vec{H} 长度为 180,且在活动分割过程中,第 1 次滑动窗口的大小 m 决定了矩阵的长度(行数),以

确保输入数据维度一致。

1.4.1 CNN

为有效提取与压缩信息,提升模型的性能和效率,本文设计了一个包含多个卷积层和池化层的 9 层 CNN 结构,如图 7 所示。在该结构中,卷积层主要负责提取空间特征;ReLU 激活函数用于增强模型的非线性表达能力;MaxPool 层则承担降维和特征选择的任务。通过这些操作,能够实现对输入数据的鲁棒性和参数共享,进而降低计算复杂度以及过拟合风险。

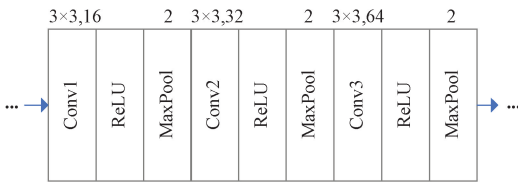


图 7 CNN 结构图

Figure 7 Convolutional neural network structure diagram

1.4.2 BiGRU

为了在处理 CSI 时捕捉长程依赖和上下文信息,提高模型对时间序列的理解能力,采用了 BiGRU 模型。BiGRU 通过双向结构同时考虑前后文信息,增强了对序列特征的捕捉能力,并且通过门控机制有效控制信息流动,减少梯度消失问题,从而提高了模型的预测准确性和鲁棒性。BiGRU 包含前向 GRU 和后向 GRU 两部分,分别将输入序列从两个方向进行处理,最后,将两个方向的处理结果进行拼接,得到最终的输出结果。BiGRU 结构如图 8 所示。

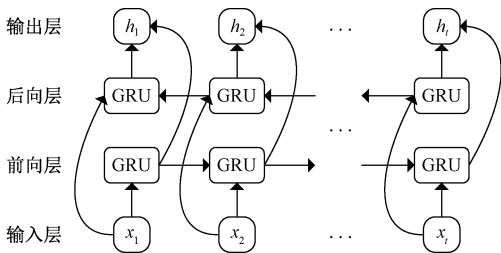


图 8 BiGRU 结构图

Figure 8 Bidirectional gated recurrent unit structure diagram

1.4.3 多头注意力机制

注意力机制能够使模型关注输入数据关键部分,并抑制无用特征的影响。MHA 机制是注意力机制的扩展,其结构如图 9 所示。通过 MHA,模型能够并行关注输入数据的不同部分,增强特征表示的多样性,提高对复杂模式的识别能力,从而提升模型的整体性能和对噪声的鲁棒性。MHA 将输入的向量 \mathbf{X} 通过 3 次不同的线性变换分别转换成 3 个维度

为 d_k 的矩阵,表示为

$$\begin{cases} \mathbf{Q} = \mathbf{X}\mathbf{W}^Q; \\ \mathbf{K} = \mathbf{X}\mathbf{W}^K; \\ \mathbf{V} = \mathbf{X}\mathbf{W}^V. \end{cases} \quad (10)$$

式中: \mathbf{Q} 为查询矩阵; \mathbf{K} 为键矩阵; \mathbf{V} 为值矩阵; \mathbf{W}^Q 、 \mathbf{W}^K 与 \mathbf{W}^V 分别为三次线性变换的权重矩阵。在每个头 i 上,使用缩放点积注意力计算公式,其计算可以表示为

$$M_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{Softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d_k}}\right) \mathbf{V}_i. \quad (11)$$

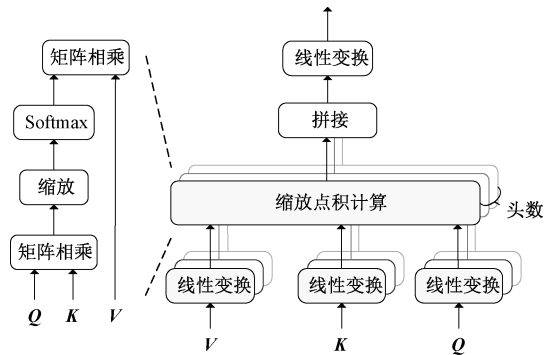


图 9 多头注意力机制结构图

Figure 9 Multi-head attention mechanism structure diagram

计算完所有头的注意力后,将它们的输出在 d_k 维度上拼接起来,再次通过一个线性变换,得到多头注意力机制的最终输出。

2 实验结果及评估

2.1 数据集描述

本文采用了两个公开数据集来验证所提模型的有效性。第 1 个数据集为 WiAR^[13],该数据集涵盖了由 10 名志愿者在会议室中执行的 16 项活动的的数据。数据采集借助两台配备 Intel 5300 NIC 的 ThinkPad T400 笔记本电脑完成,其中,发送方笔记本电脑装有 1 根天线,接收方笔记本电脑装有 3 根天线。针对每项活动,均采集了 30 个样本,采样频率设定为 30 Hz。会议室尺寸为 6 m×10 m,室内放置有一些办公家具。接收器与发射器之间相隔 4 m,志愿者所处位置距离发射器 1 m,距离接收器 3 m。WiAR 数据集中,每个活动样本所采集的数据时长超过 7 s,其中包含 2~3 s 的活动数据以及 4~6 s 的非活动数据。此外,该数据集还考虑了志愿者个体因素对识别效果的影响,特意选取了 5 名男性和 5 名女性作为志愿者参与数据采集。活动类型丰富多样,涵盖上半身活动、下半身

活动以及全身活动,上半身活动包含单臂水平挥动、双臂水平挥动、单臂高举挥动、扔纸、画勾、画叉、打电话、拍手、喝水、高抛物体;下半身活动包含前踢、侧踢;全身活动包含下蹲、坐下、弯腰、走路。

第2个数据集为BAHAR^[21],由30名志愿者在3种不同室内环境下执行的5种活动项目的数据构成。数据采集借助2台配备Intel 5300 NIC的台式电脑完成,其天线配置与WiAR数据集一致。在这3种室内环境中,E1与E2这2种环境属于视距环境,另一种环境E3为非视距环境。每种活动项目包含的活动数量为1~4个,例如“从坐到跌”项目包含3个活动,依次为坐在椅子上、跌倒和躺下。每个活动均重复进行20次,采样频率设定为320 Hz。为深入探究模型对持续变化运动的识别性能,本文将这5种活动项目作为目标域开展识别分类工作。

2.2 实验设置

本文实验基于深度学习PyTorch框架实现,以8:2的比例将数据集划分为训练集和测试集。CPU使用Intel(R) Xeon(R) Platinum 8255C, GPU使用NVIDIA RTX 2080 Ti,选用Adam优化器和CrossEntropyLoss损失函数,学习率设置为0.00001, epochs设置为100。

本文采用准确率(A)、精确率(P)、召回率(R)、F1值评估所提模型效果。计算公式如下:

$$A = \frac{TP + TN}{TP + FP + TN + FN}; \quad (12)$$

$$P = \frac{TP}{TP + FP}; \quad (13)$$

$$R = \frac{TP}{TP + FN}; \quad (14)$$

$$F1 = 2 \frac{P \cdot R}{P + R}. \quad (15)$$

式中:TP为预测为正例而且实际上也是正例;TN为预测为负例而且实际上也是负例;FP为预测为正例然而实际上却是负例;FN为预测为负例然而实际上却是正例。

2.3 系统评估

2.3.1 滑动窗口大小的选择

活动分割对人体动作识别具有重要作用,可以减少对无关特征的学习。本文基于主成分分析利用双层滑动窗口机制进行活动分割,第1次滑动窗口的大小m决定了输入数据的长度。在WiAR数据集中随机选取了6位志愿者的所有活动数据进行识别,调整不同的窗口大小,以探究滑动窗口大小对识

别的效果影响,结果见表1。在窗口大小为60时,达到了最佳效果,这是由于在每个样本中,有效部分只占25%~43%,60的滑动窗口大小占总大小的50%(样本统一压缩为120),既包含了最多活动数据43%的情况,还预留了一定容错空间,同时减少了50%的数据量。因此在后续基于WiAR数据集的实验均采用m=60。

表1 窗口大小与活动类别的影响

参数	A/%	P/%	R/%	F1值/%
m=40	93.75	94.00	93.75	93.75
m=50	95.49	95.72	95.49	95.51
m=60	96.53	96.53	96.73	96.53
m=70	95.14	95.44	95.14	95.18
m=80	95.83	96.00	95.83	95.80
n=4	100.00	100.00	100.00	100.00
n=8	98.96	98.98	98.96	98.96
n=12	96.76	96.78	96.76	96.73
n=16	96.53	96.73	96.53	96.53

2.3.2 活动多样性的影响

将WiAR数据集中的活动类别n从16个调整到4个,以探究人体动作的类别个数对识别准确率的影响,如表1所示。与预期一致,随着动作类别个数的减少,模型识别的效果显著提升,甚至在类别数为4时,达到了100%的准确率。因为类别越少,模型需要区分的特征就越少,问题的复杂性降低,模型更容易学习到每个类别的特征,从而提高了准确率。

2.3.3 距离与高度的影响

WiAR数据集还分别采集了3种不同收发设备之间距离和设备离地高度的数据,以评估其对WiFi信号的影响。数据在较为复杂的大厅环境中采集,收发设备之间的距离d=1,2,3 m,设备离地高度h为60,90,120 cm,一共有16项活动,每项动作采集10个样本。识别效果见表2。3种收发设备之间距离的识别效果都达到了100%的准确率。在设备离地高度的实验中,随着高度的增加,识别准确率也在逐步提高,120 cm的实验中达到了98.96%的准确率,这是由于不同高度可以捕捉不同的动作。60 cm高度对应于下半身,捕捉下半身活动;90 cm对应于躯干部分,感知走路等全身活动;120 cm对应于上半身,识别上半身活动。数据集对上半身和全身活动的样本更多,而下半身活动的样本较少,导致模型对下半身活动的识别效果最差。

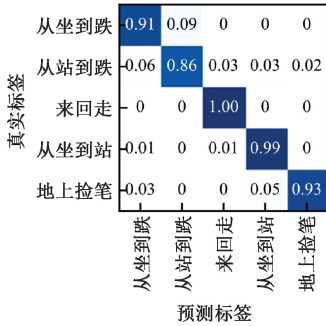
2.3.4 环境多样性的影响

除了来自WiAR数据集的视距数据之外,还评估了模型在BAHAR数据集上的性能。该数据集是

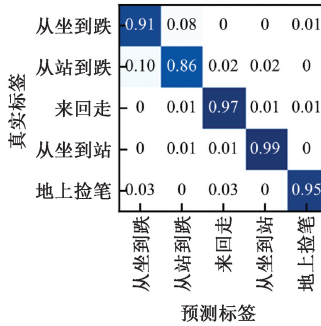
表 2 距离与高度的影响

Table 2 Effect of distance and height

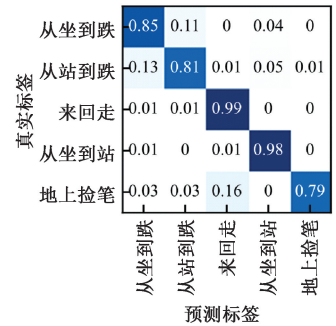
参数	A/%	F1 值/%
$d = 1 \text{ m}$	100	100
$d = 3 \text{ m}$	100	100
$d = 6 \text{ m}$	100	100
$h = 60 \text{ cm}$	94.79	94.60
$h = 90 \text{ cm}$	97.92	97.93
$h = 120 \text{ cm}$	98.96	98.91



(a) E1 视距环境混淆矩阵



(b) E2 视距环境混淆矩阵



(c) E3 非视距环境混淆矩阵

图 10 BAHAR 数据集混淆矩阵

Figure 10 BAHAR dataset confusion matrix

2.4 消融实验

2.4.1 相位的影响

在本次实验中,分别将振幅+相位和单独振幅作为数据源,经过预处理和活动分割后输入到神经网络模型中。如图 11 所示,使用振幅+相位优于单独使用振幅的识别效果,因为将振幅和相位信息相结合能提供更完整和互补的信道特征,从而提高模型的准确率。实验验证了相位的加入对人体动作识别具有重要作用。

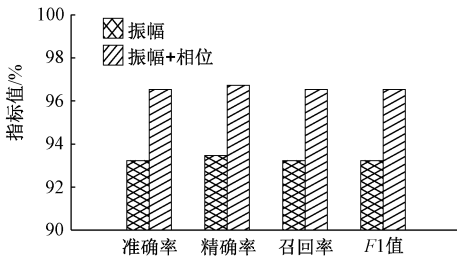


图 11 相位对识别的影响

Figure 11 Effect of phase on recognition

2.4.2 滑动窗口机制的影响

滑动窗口机制可以实现活动分割,但由于受到多径效应影响,信号会产生不规则波动,简单的滑动窗口机制无法有效筛选掉无用的信息,反而存在误将有用信息去除的可能,使得模型更难学习有效的特征。本次实验对比了使用简单滑动窗口机制与本文所提机制的性能差异。普通滑动窗口机制是直接

在 3 个不同的环境下采集的数据,探究模型在多种环境下的识别效果,包括视距与非视距环境。如图 10 所示,E1 与 E2 是视距环境,取得了不错的识别效果;E3 是非视距环境,效果相较于视距环境略差,这是由于收发设备之间存在一堵 8 cm 厚的木墙,这将大大增加识别的难度,最终达到了 91% 以上的准确率。性能比较表明,非视距环境造成的高失真对所提出模型的精度影响较小。

在全部振幅上应用主成分分析,在第一主成分上计算窗口内的标准差,最后选其最大值得到窗口索引。如图 12 所示,普通的滑动窗口机制因为误将有用信息去除,导致模型难以充分学习有效特征,性能低于不使用滑动窗口机制;本文所提滑动窗口机制可以有效去除与运动无关信息,进一步提升模型效率及性能。

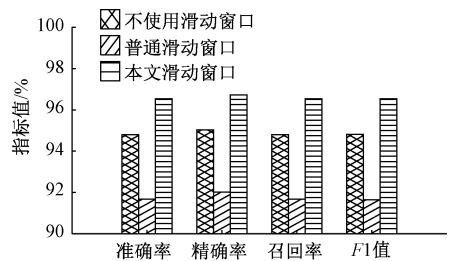


图 12 不同滑动窗口机制的效果

Figure 12 Effect of different sliding window mechanisms

2.4.3 深度学习算法的影响

为了探究本文算法模型的识别性能,本次实验对比了与其他 6 种深度学习算法的识别效果。如图 13 所示,本文算法在 WiAR 数据集上取得了最优的识别精度。

2.5 不同方法对比

表 3 比较了近 3 年同样采用 WiAR 数据集的其他方法与本文方法在识别准确率的差异,对比发现,本文方法取得了最佳的识别效果。

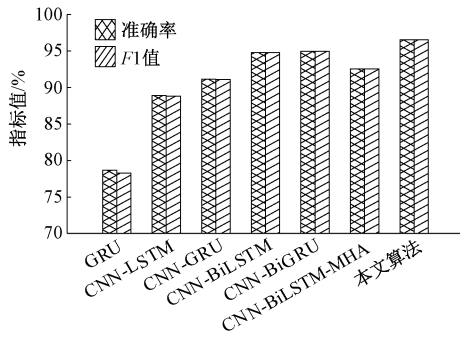


图13 不同深度学习算法的效果

Figure 13 Effect of different deep learning algorithms

表3 不同方法的效果对比

Table 3 Effect comparison of different methods

方法	核心算法	A/%
PCWCNN ^[22]	Wavelet CNN	95.50
DeepFocus ^[23]	CSAM-BiLSTM	94.62
LCED ^[24]	LSTM-CNN	95.00
本文方法	CNN-BiGRU-MHA	96.53

2.6 复杂度分析

本文对所提模型进行了全面的复杂度分析。模型包含 69.9 MB 可训练参数,由于需要处理序列信息和注意力计算,总浮点运算次数达 177.46 GFLOPs。在推理性能方面,模型平均推理时间仅为 1.84 ms(标准差 0.60 ms),处理速度达到 543.20 帧/s,远超实时处理所需的 30 帧/s,充分满足了实际应用需求。

3 结论

本文提出了一种基于 CSI 主成分的双层滑动窗口机制的人体动作识别方法。利用基于主成分分析的双层滑动窗口机制对预处理后的 CSI 数据进行活动分割,有效去除与运动无关的信息,提升模型的训练效率;为了充分挖掘数据潜在特征之间的关系,采用 CNN-BiGRU 架构对 CSI 数据的空间和时间两个维度进行分析,并融合多头注意力机制专注关键信息部分,实现对人体动作的高精度识别。本文对两个公开数据集进行了广泛的实验,包括滑动窗口大小选择、活动类别个数、视距与非视距环境等条件对识别的影响,并且对比了所提方法与几种现有方法的识别性能。实验结果表明,所提方法在两个公开数据集上可以大幅去除无关信息,并取得了良好的识别效果,优于其他现有的识别方法。

未来将进一步提高所提方法对环境变化的抵抗能力和对未知动作的识别性能,鉴于基于 CSI 的人体动作识别可能对用户隐私造成威胁,后续研究将

着重考虑保障用户隐私安全。

参考文献:

- [1] MOSLEH S, CODER J B, SCULLY C G, et al. Monitoring respiratory motion with Wi-Fi CSI: characterizing performance and the breathe smart algorithm[J]. IEEE Access, 2022, 10: 131932-131951.
- [2] TANG Z L, LIU Q Q, WU M J, et al. Wi-Fi CSI gesture recognition based on parallel LSTM-FCN deep space-time neural network[J]. China Communications, 2021, 18(3): 205-215.
- [3] HAMEED H, USMAN M, TAHIR A, et al. Pushing the limits of remote RF sensing by reading lips under the face mask[J]. Nature Communications, 2022, 13: 5168.
- [4] SHALABY E, ELSHENNAWY N, SARHAN A. Utilizing deep learning models in CSI-based human activity recognition[J]. Neural Computing and Applications, 2022, 34(8): 5993-6010.
- [5] WANG J C, TIAN Z S, ZHOU M, et al. Leveraging hypothesis testing for CSI based passive human intrusion direction detection[J]. IEEE Transactions on Vehicular Technology, 2021, 70(8): 7749-7763.
- [6] BRENA R F, ESCUDERO E, VARGAS-ROSALES C, et al. Device-free crowd counting using multi-link Wi-Fi CSI descriptors in Doppler spectrum[J]. Electronics, 2021, 10(3): 315.
- [7] SOTO J C H, GALDINO I, CABALLERO E, et al. A survey on vital signs monitoring based on Wi-Fi CSI data[J]. Computer Communications, 2022, 195: 99-110.
- [8] KOTARU M, KATTI S. Position tracking for virtual reality using commodity Wi-Fi[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2671-2681.
- [9] XU Q Y, HAN Y, WANG B B, et al. Indoor events monitoring using channel state information time series[J]. IEEE Internet of Things Journal, 2019, 6(3): 4977-4990.
- [10] GU Y, REN F J, LI J. PAWS: passive human activity recognition based on Wi-Fi ambient signals[J]. IEEE Internet of Things Journal, 2016, 3(5): 796-805.
- [11] HALPERIN D, HU W J, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information[J]. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 53.
- [12] LI F M, AL-QANESS M, ZHANG Y, et al. A robust and device-free system for the recognition and classification of elderly activities[J]. Sensors, 2016, 16(12): 2043.
- [13] GUO L L, WANG L, LIN C, et al. Wiar: a public dataset for Wi-Fi-based activity recognition[J]. IEEE Ac-

- cess, 2019, 7: 154935–154945.
- [14] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [15] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2013: 6645–6649.
- [16] MOSHIRI P F, SHAHBAZIAN R, NABATI M, et al. A CSI-based human activity recognition using deep learning[J]. Sensors, 2021, 21(21): 7225.
- [17] SHI Z G, ZHANG J A, XU R, et al. Deep learning networks for human activity recognition with CSI correlation feature extraction[C]//ICC2019-2019 IEEE International Conference on Communications (ICC). Piscataway: IEEE, 2019: 1–6.
- [18] SHANG S N, LUO Q Y, ZHAO J J, et al. LSTM-CNN network for human activity recognition using Wi-Fi CSI data[J]. Journal of Physics: Conference Series, 2021, 1883(1): 012139.
- [19] YAMAK P T, LIY J, GADOSEY P K, et al. A comparison between ARIMA, LSTM, and GRU for time series forecasting[C]//Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence. New York: ACM, 2020: 49–55.
- [20] MING X, CHENG W, ZHU R L, et al. Human activities recognition with amplitude-phase of channel state information using deep residual networks[C]//2022 IEEE 17th Conference on Industrial Electronics and Applications (ICIEA). Piscataway: IEEE, 2022: 1–6.
- [21] ALSAIFY B A, ALMAZARI M M, ALAZRAI R, et al. A dataset for Wi-Fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments[J]. Data in Brief, 2020, 33: 106534.
- [22] SHOWMIK I A, SANAM T F, IMTIAZ H. Human activity recognition from Wi-Fi CSI data using principal component-based wavelet CNN[J]. Digital Signal Processing, 2023, 138: 104056.
- [23] ZHOU Q Z, XING J C, YANG Q L, et al. Measuring intrinsic human activity information using Wi-Fi-based attention model[J]. Measurement, 2022, 195: 111084.
- [24] GUO L L, ZHANG H, WANG C, et al. Towards CSI-based diversity activity recognition via LSTM-CNN encoder-decoder neural network[J]. Neurocomputing, 2021, 444: 260–273.

Human Activity Recognition Method Based on CSI Principal Component Segmentation

RAO Zhuang¹, DING Dazhao², WANG Yijing²

(1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China; 2. Songshan Laboratory, Zhengzhou 450000, China)

Abstract: The traditional method of human activity recognition based on channel state information (CSI) suffers from issues such as input data redundancy and limited feature extraction. To address this, a human activity recognition approach based on CSI principal components and a dual-layer sliding window mechanism was proposed. Firstly, outlier removal and noise reduction were performed on the amplitude the use of a dual-layer sliding window mechanism based on principal component analysis enabled activity segmentation of preprocessed CSI data to eliminate irrelevant information and enhance model training efficiency. Subsequently, spatial and temporal analysis of the CSI data was conducted using convolutional neural network and bidirectional gated recurrent unit, with the integration of a multi-head attention mechanism to focus on key information for achieving high-precision recognition of human activities. Experimental validation was performed using the WiAR and BAHAR public datasets, demonstrating that the proposed method could effectively recognize various human activities in diverse environments, while reducing the data volume by 5%. The accuracy achieved on the WiAR dataset was 96.53%, indicating superior performance compared to existing methods.

Keywords: channel state information; activity segmentation; convolutional neural network; bidirectional gated recurrent unit; multi-head attention mechanism